DOCUMENT RESUME

ED 132 204 TM 005 941

AUTHOR Naccarato, Richard W.; And Others
TITLE Student Ratings Halo Effect Reduction by Alteration of Standard Instructions. 76-18. Educational Assessment Center Project: 554.
INSTITUTION Washington Univ., Seattle. Educational Assessment Center.
PUB DATE Nov 76
NOTE 21p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS College Students; Correlation; *Course Evaluation; Educational Diagnosis; Effective Teaching; *Higher Education; Item Analysis; *Rating Scales; *Student Evaluation of Teacher Performance; Testing; *Testing Problems
IDENTIFIERS *Halo Effect; *Instructional Assessment System Form B

ABSTRACT

The purpose of this study was to determine if correlations among student ratings items designed to be diagnostic could be lowered through use of special instructions to raters. The authors argue that the lowering of inter-item correlations is indicative of a reduction of the halo effect which leads to greater item diagnosticity. The experimental group first ranked items in terms of importance, then rated the course with the diagnostic items, then rated the course with the general items. This order was reversed for the control group. The correlations among items were significantly lower for the experimental group. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

EAC REPORTS

EDUCATIONAL ASSESSMENT CENTER

University of Washington, 1400 Campus Parkway, PB–30, Seattle, Washington 98195 (206) 543–1170

Instructional Evaluation

Educational, Psychological, and Survey Research

Test Administration, Construction, Scoring, and Analysis

Educational Assessment Center

University of Washington

November 1976

76-19

Student Ratings Halo Effect Reduction by Alteration of Standard Instructions

Richard W. Naccarato, Miriam A. Goldstein &

Gerald M. Gillmore

Abstract

The purpose of this study was to determine if correlations among student ratings items designed to be diagnostic could be lowered through use of special instructions to raters. The authors argue that the lowering of inter-item correlations is indicative of a reduction of the halo effect which leads to greater item diagnosticity. The experimental group first ranked items in terms of importance, then rated the course with the diagnostic items, then rated the course with the general items. This order was reversed for the control group. The correlations among items were significantly lower for the experimental group.

Educational Assessment Center Project: 554

Student Ratings of Halo Effect Reduction by Alteration of Standard Instructions

Richard W. Naccarato, Miriam A. Goldstein, and Gerald M. Gillmore

Students' evaluations of college courses continue to receive attention from faculty and administration in institutions of higher learning. In a recent paper on student-faculty evaluations Permut speaks of the demand for "accountability in higher education" and of the brighter spotlight being shone upon student evaluations, not only by administrators and faculty, but by students and governmental agencies (1974, p. 41). Whether student ratings are to be used for administrative decision-making or instructional improvement, it is desirable to reduce the effect of extraneous factors on the results. This study centers itself around the "halo" effect that apparently exists in many rating situations, and assesses the impact of a strategy to reduce this contaminating effect upon the usefulness of student ratings for diagnosis of instructional problems.

Costin, Greenough, and Menges (1971) defined the halo effect as the tendency of raters to respond similarly to all items on the basis of some set impression. The origin or causes of these set impressions is relatively unknown; however, most studies have attributed the impressions to various perceptual and attitudinal processes within the individual. Widlak, McDaniel, and Feldhusen (Note 3) performed a factor analysis of student ratings results in order to assess existing halo effects. Using the Course-Instructor Evaluation (CIE) from Purdue University, they correlated 18 evaluation items and concluded that the halo effect was so strong in the CIE that the specific item ratings may have little diagnostic value in assessing a teacher's strengths and weaknesses. In a

statistical analysis of data from the first year's use of the Instructional Assessment System (IAS) at the University of Washington (involving the instrument used in this study), Gillmore (Note 1) reported fairly high correlations among items designed to be diagnostic in purpose. The correlations, computed with classes as the unit, averaged about 70. Gillmore suggested that these high correlations could indicate the presence of a strong halo effect, and importantly may limit the diagnostic value of the items. Gillmore cautioned that "...one who does well in his teaching in one area [possibly] also tends to do well in other areas, and vice versa. In other words, the halo may be, in fact, an accurate perception." (21-22).

One very evident way in which student rating results can be used to improve instruction is for the instructor to concentrate on those items on which he is rated low, and try to improve in the areas assessed by the items. In other words he can use items on which he is rated low as diagnostic of particular problems. However, insofar as items are highly correlated across classes, the lower rated items will not be indicative of particular problems. The relatively low mean values will be a result of random error or be an artifact of the intensity with which the item is worded. Thus, high inter-item correlations restrict the diagnostic value of the instrument, whether the high correlations accurately reflect reality or not.

Thus far, we have based our arguments, both for the existence of a halo effect and for the consequent loss of item diagnosticity, on high inter-item correlations across classes. However, halo effects are usually thought of as emanating from an individual rather than a group. Clearly for student instructional ratings, a halo effect must be operating within individuals in order to be operating for classes. High inter-item correlations across individuals

within a class would seem to be necessary if not sufficient evidence of the existence of a halo effect at the individual level. Furthermore, to reduce correlations among item means across classes, which are caused by a halo effect, one must be able to reduce the inter-item correlations within classes.

The purpose of the present study is to determine if the correlations among the diagnostic items of the IAS can be reduced by altering the standard procedure for administering the forms. Specifically, standard administrative procedure was altered in two regards. First, IAS forms contain items within three sections, with items within the initial section being designed to be global or general in nature. Since students normally complete this section prior to continuing on to the diagnostic items, the general items may produce a set to respond at a given level throughout the instrument. This level would probably be based on the students overall judgment of the quality of the course and instructor. Thus, our first strategy for reducing inter-item correlations was to have experimental subjects respond to the diagnostic portion of the form prior to responding to the general items.

Our second strategy was based on the notion that students possibly do not take the time and effort to read and consider items carefully before responding, and, hence, do not make careful discriminations based on item content. To counteract this tendency, if it exists, we forced experimental students to make fine discriminations among the diagnostic items by requesting that they be ranked in terms of importance in assessing teaching effectiveness prior to being used to rate the course. As a somewhat serendipitous result of this strategy we were also able to obtain data on the relative importance of the various items as perceived by students.

Method

Subjects. Ninety-six students from an elementary economics course at the University of Washington participated in the study. Four quiz sections were randomly selected from the twenty sections comprising the course. Two of these sections were randomly chosen from the four and combined into an experimental group, (N = 49). The remaining two sections were combined to form the control group, (N = 47). The separate quiz sections met twice a week, whereas the entire group attended lectures three times per week. The evaluation instrument was administered to the four quiz sections separately at their weekly meeting.

Instrument. IAS form B (Gillmore, Note 2) was administered to experimental and control groups (see Appendix A for complete form).

Procedure. Permission from the course instructor had been secured prior to visiting the quiz sections and the teaching assistants (TA's) for the sections were aware that their section might be chosen that day for participation in the experiment.

When the experimenter arrived at the classrooms, the TA left the room. The tailored instructions (See Appendix C for complete instructions) were read aloud to the two sections comprising the experimental group. These students were instructed to bypass the demographic items and the four global items, to rank order separately the remaining eighteen diagnostic instructor feedback and course information items, and then to grid in their evaluative responses to these diagnostic items before responding to the former items. Standard instructions (See Appendix B for complete instructions) were given to the two sections composing the control group, in which students responded to all evaluation items in the order in which they occurred. Subsequently Ss were asked to rank order the diagnostic instructor and course items as to their importance

in measuring teaching effectiveness. Both control and experimental groups were told that they were rating the main instructor for the course and not the TA for their section.

Results and Conclusions

The primary research hypothesis of this study concerned itself with the reduction of the halo effect as evidenced by high inter-item correlations. For our evaluative instrument the diagnostic items of interest are the instructor feedback and student information items of Table 1. The resulting inter-item correlations for these two sets of items, under the tailored instructions given to the experimental group and the standard instructions given to the control group, appear in Table 2. Italisized correlations represent those of the experimental group.

Insert Tables 1 and 2 about here

The inter-item correlations among items 5 through 15 and 16 through 22 tend to be smaller for the experimental group than those of the control group condition (Table 2). The average inter-item correlation ($\bar{r}_{ij}$) within both instructor feedback and student information items for the experimental group was .29, whereas for the control group $\bar{r}_{ij}$ equalled .43 for the instructor feedback and .46 for the course information items. To test for pairwise directional differences between experimental and control group correlations a sign test (Winer, 1971) was performed on the pairs of correlations in Table 2. Of the 55 pairs of correlation coefficients within the diagnostic instructor feedback section, 38 of the experimental group correlations were less than those of the control group, a difference significant at the .01 level. The result of the sign test for the student information items showed 16 of the 21 pairs of

correlations for the experimental group to be less than those of the control group - a difference significant at the .05 level. We can conclude from these results, then, that the tailored instructions given to the experimental group resulted in reduced inter-item correlations among the two sets of diagnostic evaluation items.

Additional evidence exists to show that the students within the experimental group continued to show more discrimination among items between the instructor feedback section and course information section, as well as within these evaluative sections. Total ratings were computed for the eleven instructor feedback items and the seven course information items for both experimental and control groups. The correlation between total instructor feedback and course information sections across all students within the experimental group equalled .53, whereas the same correlation for the control group was .71. These results may be taken as further support for the contention that the experimental instructions cause the student to look more discerningly at the specific items rather than to be affected by some overriding attitude, or halo effect, throughout the evaluation. A z-test for differences between these correlations did show, however, no statistically significant difference between the groups.

It is interesting to ask if the experimental treatment altered the item means in comparison to the control group. t-tests were performed between experimental and control group mean responses on all diagnostic items within the instructor feedback and student information sections. No obtained t value between group mean responses reached significance at the .05 level. Furthermore, the experimental group gave more favorable ratings on ten items, and less favorable ratings on eight items. This difference is not significant. Thus, there is no evidence that the experimental treatment altered the overall level at which students responded.

Importance Rankings

Instructions to both the experimental and control group students included having each student rank the items within each section in terms of importance. The only difference between groups was that the experimental group ranked the items prior to using them to rate the course and instructor, the control group did their rankings subsequent to their ratings. The median rank of each item for both groups is found in Table 3. Also found in Table 3 are the relative ranks of the items in terms of these medians.

Insert Table 3 about here

In general, there was a high degree of agreement between experimental and control group members in terms of the relative importance of items. The rank correlation between the ranks for the instructor feedback items was .79, the same correlation for the student information items was a perfect 1.00.

Within the instructor feedback section, the highest ranked items were those dealing with the instructor's explanations and organization. Items dealing with characteristics of the instructor, e.g., his/her enthusiasm, interest, clarity of objectives, and availability of extra help were rated as less important. Within the student information section, amount learned in the course was rated most highly followed by the relevance and usefulness of the course content. Instructor interest in student learning and use of class time were intermediately ranked. Grading, clarity of responsibilities, and assigned work were rated as least important. From a pedagogical point of view, the rankings by students seems very sound. However, it should be kept in mind that these rankings were applied to a specific course, not courses in general.

Discussion

The primary purpose of this study was to explore whether correlations among diagnostic items of a student ratings form could be reduced through using special instructions to raters. These instructions differed from standard instructions in two ways: Students rank-ordered items in terms of importance prior to using them to rate the course, and students responded to the diagnostic items prior to responding to general evaluational items.

The special instructions were successful in reducing the inter-item correlations relative to the same correlations deriving from the ratings of a group using standard instructions. We theorized that this reduction could be indicative of greater diagnostic value of the ratings of the experimental group. This implication is clearly based on an indirect and statistical agreement, but reduced correlations among items within a class are not sufficient to claim greater diagnosticity of those items. In the extreme case, inter-item correlations can be reduced by including irrelevant and poorly-written items on the form; a method which would clearly reduce diagnosticity. Further studies should be conducted in which the methodology of this study is combined with systematic manipulations of some specific teaching behaviors, e.g., poor vs good explanations, while holding others constant. Studies of this sort could more directly confront the issue of relative item diagnosticity.

Further study is also necessary to assess the relative importance of the two strategies used in this study to influence correlations among items. Having students rank items before using them for rating the course was confounded with having students respond to the diagnostic items prior to responding to the general items. It is presently impossible to determine which of these strategies is effective, or whether it is a combination of the two. At least

two additional groups should be assessed; one for which standard instructions are only modified to include prior ranking of items and one for which standard instructions are only modified to include responding to diagnostic items first.

To conclude, the basic purpose of this study was achieved; that is, non-standard instructions were developed which successfully reduced correlations among items. We feel these lowered correlations may reflect an increase in the information arising from these items specifically for the diagnosis of instructional problems. Further research, more direct in nature, is needed to validate our assumption.

Table 1

Items Within the Instructor Feedback and the Student Information Sections (Form B) of IAS

INSTRUCTOR FEEDBACK ITEMS

5. Course organization was:
6. Sequential presentation of concepts was:
7. Explanations by instructor were:
8. Instructor's ability to present alternative explanations when needed was:
9. Instructor's use of examples and illustrations was:
10. Instructor's enhancement of student interest in the material was:
11. Student confidence in instructor's knowledge was:
12. Instructor's enthusiasm was:
13. Clarity of course objectives was:
14. Interest level of class sessions was:
15. Availability of extra help when needed was:

STUDENT INFORMATION ITEMS

16. Use of class time was:
17. Instructor's interest in whether students learned was:
18. Amount you learned in the course was:
19. Relevance and usefulness of course content is:
20. Evaluative and grading techniques (tests, papers, projects, etc.) were:
21. Reasonableness of assigned work was:
22. Clarity of student responsibilities and requirements was:

Table 2

Inter-Item Correlations of Instructor Feedback Items and Student Information Items in Experimental and Control Groups[1]

EM	6	7	8	9	10	11	12	13	14	15	17	18	19	20	21	22
5	*.48* .78	*.32* .28	*.26* .46	*.19* .33	*.31* .30	*.17* .31	*-.05* .56	*.58* .21	*-.06* .08	*-.11* .32						
6		*.56* .43	*.29* .51	*.38* .45	*.30* .40	*.19* .37	*.24* .49	*.54* .35	*.004* .42	*.04* .28						
7			*.29* .69	*.58* .46	*.38* .42	*.36* .27	*.31* .38	*.36* .24	*.33* .27	*.14* .20						
8				*.42* .59	*.14* .49	*.37* .44	*.16* .41	*.19* .38	*.30* .56	*.16* .39						
9					*.44* .45	*.57* .44	*.31* .33	*.21* .36	*.25* .40	*.37* .26						
0						*.30* .54	*.40* .46	*.20* .28	*.62* .68	*.20* .36						
1							*.44* .36	*.04* .18	*.37* .45	*.45* .42						
2								*.03* .61	*.46* .38	*.50* .47						
3									*-.16* .38	*.35* .19						
4										*.33* .19						
6											*.34* .29	*-.006* .37	*-.02* .26	*.41* .16	*.23* .17	*.29* .30
7												*.36* .56	*.10* .44	*.41* .39	*.41* .35	*.46* .68
8													*.43* .68	*.33* .50	*.45* .62	*.10* .61
9														*.20* .42	*.24* .64	*.12* .50
0															*.44* .63	*.36* .48
1																*.47* .62

[1] Italic numbers represent the experimental condition. Items 5-15 are the instructor feedback items, while items 16-22 are the student information items.

Table 3

Median Importance Ranks and Relative Rank of

Items in Experimental and Control Conditions

Item	Median Rank		Relative Rank	
Instructor Feedback Items	Exper.	Control	Exper.	Control
5	4.7	5.6	2	4
6	5.6	6.4	5	7
7	2.4	1.7	1	1
8	5.4	5.4	4	3
9	5.0	4.0	3	2
10	7.1	5.4	8	5
11	7.3	6.1	9	6
12	6.7	7.1	7	8
13	6.5	7.6	6	10
14	8.6	7.6	11	10
15	7.9	7.6	10	10
Student Information Items				
16	4.2	4.2	4	4
17	3.5	3.3	3	3
18	1.8	1.8	1	1
19	2.3	2.2	2	2
20	4.9	4.9	5	5
21	5.4	5.6	7	7
22	5.2	5.3	6	6

Reference Notes

1. Gillmore, G. M. Statistical analyses of the data from the first year of use of the student ratings forms of the University of Washington Instructional Assessment System. (EAC Report 76-9). Seattle, Washington: University of Washington, Educational Assessment Center, 1975.

2. Gillmore, G. M. A brief description of the University of Washington Instructional Assessment System. (EAC Report 276). Seattle, Washington: University of Washington, Educational Assessment Center, 1974.

3. Widlak, F. W., McDaniel, E. D.,& Feldhusen, J. F. Factor analysis of an instructor rating scale. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 1973.

References

Costin, F., Greenough, W. T., & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

Permut, S. E. Cue utilization patterns in student-faculty evaluation. Journal of Psychology, 1973, 83, 41-48.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

Appendix A: Form B, Instructional Assessment System, Educational Assessment Center, University of Washington.

Appendix B: Regular Control Group Instructions.

Appendix C: Tailored Instructions.

INSTRUCTIONAL ASSESSMENT SYSTEM

INSTRUCTOR _____

COURSE _____ SECTION _____

EDUCATIONAL ASSESSMENT CEN
UNIVERSITY OF WASHINGTON

DIRECTIONS: YOU MAY RETURN THIS QUESTIONNAIRE COMPLETELY OR PARTIALLY UNANSWERED WITHOUT PENALTY. USE A NO. 2 PENCIL AND MAKE MARKS FIRM BUT NOT SHINY. DO NOT CROSS OUT GRIDS. DO NOT MAKE STRAY MARKS. ERASE CLEANLY IF YOU CHANGE AN ANSWER.

WHEN REGISTERING, WAS THIS A COURSE YOU WANTED TO TAKE?	YES	NO	NEUTRAL			
IS THIS COURSE:	IN YOUR MAJOR	IN YOUR MINOR OR PROGRAM REQUIREMENT	A DISTRIBUTION REQUIREMENT	AN ELECTIVE	OTHER	
YOUR CLASS:	FRESHMAN	SOPHOMORE	JUNIOR	SENIOR	GRADUATE	OTHER
GRADE YOU EXPECT TO RECEIVE:	A	B	C	D	E	PASS

STUDENT RATING FORM

B

IMPORTANT: IN RATING THIS COURSE, RESPOND TO EACH ITEM CAREFULLY AND THOUGHTFULLY. AVOID LETTING YOUR RESPONSES TO SOME ITEMS INFLUENCE YOUR RESPONSES TO OTHERS. KEEP THE PURPOSE OF EACH SECTION IN MIND AS YOU RATE THE COURSE.

E—EXCELLENT
VG—VERY GOOD
G—GOOD
F—FAIR
P—POOR
VP—VERY POOR

SECTION 1:
TO PROVIDE A GENERAL EVALUATION.

1. THE COURSE AS A WHOLE WAS: 1.
2. THE COURSE CONTENT WAS: 2.
3. THE INSTRUCTOR'S CONTRIBUTION TO THE COURSE WAS: 3.
4. THE INSTRUCTOR'S EFFECTIVENESS IN TEACHING THE SUBJECT MATTER WAS: 4.

SECTION 2:
TO PROVIDE DIAGNOSTIC FEEDBACK TO THE INSTRUCTOR.

5. COURSE ORGANIZATION WAS: 5.
6. SEQUENTIAL PRESENTATION OF CONCEPTS WAS: 6.
7. EXPLANATIONS BY INSTRUCTOR WERE: 7.
8. INSTRUCTOR'S ABILITY TO PRESENT ALTERNATIVE EXPLANATIONS WHEN NEEDED WAS: 8.
9. INSTRUCTOR'S USE OF EXAMPLES AND ILLUSTRATIONS WAS: 9.
10. INSTRUCTOR'S ENHANCEMENT OF STUDENT INTEREST IN THE MATERIAL WAS: 10.
11. STUDENT CONFIDENCE IN INSTRUCTOR'S KNOWLEDGE WAS: 11.
12. INSTRUCTOR'S ENTHUSIASM WAS: 12.
13. CLARITY OF COURSE OBJECTIVES WAS: 13.
14. INTEREST LEVEL OF CLASS SESSIONS WAS: 14.
15. AVAILABILITY OF EXTRA HELP WHEN NEEDED WAS: 15.

SECTION 3:
TO PROVIDE INFORMATION ABOUT THE COURSE TO OTHER STUDENTS.

16. USE OF CLASS TIME WAS: 16.
17. INSTRUCTOR'S INTEREST IN WHETHER STUDENTS LEARNED WAS: 17.
18. AMOUNT YOU LEARNED IN THE COURSE WAS: 18.
19. RELEVANCE AND USEFULNESS OF COURSE CONTENT IS: 19.
20. EVALUATIVE AND GRADING TECHNIQUES (TESTS, PAPERS, PROJECTS, ETC.) WERE: 20.
21. REASONABLENESS OF ASSIGNED WORK WAS: 21.
22. CLARITY OF STUDENT RESPONSIBILITIES AND REQUIREMENTS WAS: 22.

SECTION 4:
OPTIONAL ITEMS—USE ONLY AS DIRECTED.

27. 28. 29. 30. 23. 24. 25. 26.

© University of Washington 1974

Appendix B: Regular Control Group Instructions

Hello, I'm _____ from the Educational Assessment Center and I'm doing a study to learn more about how students rate their courses. I'd like you to respond to this questionnaire. While this is not a regular end-of-the-quarter rating, the results will be given to the instructor after the course is over. If items refer to the instructor, rate your professor and not your T.A. Please respond to every question. Does anyone need a pencil?

Beginning at the top of the questionnaire, where you are asked for information about yourself, please respond to the entire questionnaire. I'll wait. (Wait.) Now, let's go back to Section II. Rank order all of the 11 items from 1 to 11 judging what you believe are most important as feedback items to the instructor's teaching effectiveness. Remember, 1 is most important and 11 is least important. Place your ranks to the left of the printed item number. Do not go back and change your responses. Do the same for the 7 items in Section III, ranking them from 1 to 7. Again, 1 is most important and 7 is least important.

Are there any questions?

(When finished, thank the students.)

Appendix C: Tailored Instructions

Hello, I'm _____from the Educational Assessment Center and I'm doing a study to learn more about how students rate their courses. I'd like you to respond to this questionnaire. While this is not a regular end-of-quarter rating, the results will be given to the instructor after the course is over. If items refer to the instructor, rate your professor and not your T.A. Please respond to every question. I'm going to pass out questionnaires. Please leave them face-down until I give you further instructions. Does anyone need a pencil?

I would like to begin with Section II. Read the items - there are 11 of them. Rank order all of the items in Section II from 1 to 11 judging how important they are as feedback items to the instructor's teaching effectiveness. 1 is most important and 11 is least important. Place your ranks to the left of the printed item number. (Pause) Go back and grid in the items in the order in which you ranked them...one first, and so on. Do the same for Section III, ranking the 7 items from 1 to 7. Again, 1 is most important and 7 is least important.

When you have completed Section III, go to the top of the questionnaire where you are asked for information about yourself. Please respond. Then go to Section I. Do not rank order these items. Simply respond to the choices.

Are there any questions?

(When finished, thank the students.)